# 

# **Supplementary information**

**for**

**Temperature and nutrient gradients correspond with lineage-specific microdiversification in the ubiquitous and abundant *Limnohabitans* freshwater genus**

Ruben Props1,2 and Vincent J. Denef2, ¥

1 Center for Microbial Ecology and Technology (CMET), Ghent University, Coupure Links 653, B-9000 Gent, Belgium

2 Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, US

¥ Corresponding author: 1141 Kraus Natural Science, 830 N. University, Ann Arbor, MI 48109 vdenef@umich.edu, Phone: +1 (734) 764 6481, Fax: +1 (734) 763 0544

**Supplementary Table S1**: Results from hypergeometric testing for gene set enrichment in accessory genomes of putative Limnohabitans MAGs. Only KEGG subsystem categories with an adjusted p-value less than 0.05 are shown. Description: KEGG subsystem category that was tested. GeneRatio: number of genes in KEGG subsystem category relative to the total number of annotated genes in accessory genome. BgRatio: number of genes in KEGG substem category relative to the total number of annotated genes in complete MAG. P.adjust: Benjamini-Hochberg adjusted p-value, qvalue: false discovery rate. Genome: MAG for which enrichment was tested.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Description** | **GeneRatio** | **BgRatio** | **pvalue** | **p.adjust** | **qvalue** | **Genome** |
| ABC transporters | 28/184 | 122/1531 | 0.000292 | 0.00446 | 0.003848 | MAG4.FA-M110-DN |
| ABC transporters | 15/50 | 108/1357 | 2.68E-06 | 4.96E-05 | 4.09E-05 | MAG8.SU-M110-DCMD |
| Aminobenzoate degradation | 5/50 | 16/1625 | 7.62E-05 | 0.002363 | 0.002086 | MAG9.SU-M15-SN |
| Arginine and proline metabolism | 4/32 | 19/1515 | 0.00051 | 0.010702 | 0.009656 | MAG1.FA-MLB-DN |
| Arginine and proline metabolism | 12/375 | 24/1886 | 0.000853 | 0.013655 | 0.012577 | MAG2.FA-MLB-SN |
| Bacterial motility proteins | 29/375 | 65/1886 | 3.54E-06 | 9.43E-05 | 8.69E-05 | MAG2.FA-MLB-SN |
| Bacterial motility proteins | 14/189 | 50/1685 | 0.000744 | 0.027167 | 0.025071 | MAG3.FA-MLB-SN |
| Bacterial motility proteins | 13/184 | 44/1531 | 0.001264 | 0.015422 | 0.013306 | MAG4.FA-M110-DN |
| Bacterial motility proteins | 7/104 | 18/1087 | 0.00079 | 0.016189 | 0.014963 | MAG5.SP-M110-DD |
| Bacterial motility proteins | 4/19 | 44/1085 | 0.005849 | 0.019168 | 0.01153 | MAG7.SU-MLB-SD |
| Bacterial motility proteins | 7/50 | 30/1357 | 6.40E-05 | 0.00079 | 0.000652 | MAG8.SU-M110-DCMD |
| Bacterial secretion system | 5/33 | 23/1321 | 0.000174 | 0.001507 | 0.001403 | MAG6.SP-M15-SD |
| Biofilm formation | 12/375 | 19/1886 | 4.22E-05 | 0.000844 | 0.000777 | MAG2.FA-MLB-SN |
| Biofilm formation | 6/33 | 12/1321 | 1.26E-07 | 1.64E-06 | 1.53E-06 | MAG6.SP-M15-SD |
| Flagellar assembly | 3/19 | 25/1085 | 0.008215 | 0.019168 | 0.01153 | MAG7.SU-MLB-SD |
| Function unknown | 16/189 | 65/1685 | 0.001469 | 0.035748 | 0.03299 | MAG3.FA-MLB-SN |
| Membrane and intracellular structural molecules | 7/104 | 11/1087 | 1.45E-05 | 0.000594 | 0.000549 | MAG5.SP-M110-DD |
| Prokaryotic Defense System | 6/47 | 36/1682 | 0.000355 | 0.015271 | 0.013458 | MAG10.SU-M15-SN |
| Quorum sensing | 8/32 | 113/1515 | 0.001727 | 0.018136 | 0.016363 | MAG1.FA-MLB-DN |
| Quorum sensing | 29/184 | 97/1531 | 8.00E-07 | 2.44E-05 | 2.10E-05 | MAG4.FA-M110-DN |
| Quorum sensing | 11/50 | 92/1357 | 0.000315 | 0.002329 | 0.001921 | MAG8.SU-M110-DCMD |
| Secretion system | 32/375 | 67/1886 | 1.52E-07 | 6.07E-06 | 5.59E-06 | MAG2.FA-MLB-SN |
| Secretion system | 23/189 | 69/1685 | 4.56E-07 | 3.33E-05 | 3.07E-05 | MAG3.FA-MLB-SN |
| Secretion system | 19/184 | 55/1531 | 7.51E-06 | 0.000153 | 0.000132 | MAG4.FA-M110-DN |
| Secretion system | 13/33 | 54/1321 | 6.21E-11 | 1.62E-09 | 1.50E-09 | MAG6.SP-M15-SD |
| Transporters | 95/375 | 297/1886 | 4.14E-08 | 3.31E-06 | 3.05E-06 | MAG2.FA-MLB-SN |
| Transporters | 55/184 | 226/1531 | 1.25E-08 | 7.62E-07 | 6.57E-07 | MAG4.FA-M110-DN |
| Transporters | 8/19 | 184/1085 | 0.008205 | 0.019168 | 0.01153 | MAG7.SU-MLB-SD |
| Transporters | 25/50 | 224/1357 | 2.20E-08 | 8.14E-07 | 6.72E-07 | MAG8.SU-M110-DCMD |
| Two-component system | 8/50 | 45/1357 | 0.000146 | 0.001349 | 0.001113 | MAG8.SU-M110-DCMD |

**Supplementary Table S2**: Blastx results of major chemotaxis genes found in MAG8.SU-M110-DCMD to the non-redundant protein database and blastp search results specifically to the Limnohabitans sp. 63ED37-2 genome and MAG7 (LimC genomes).

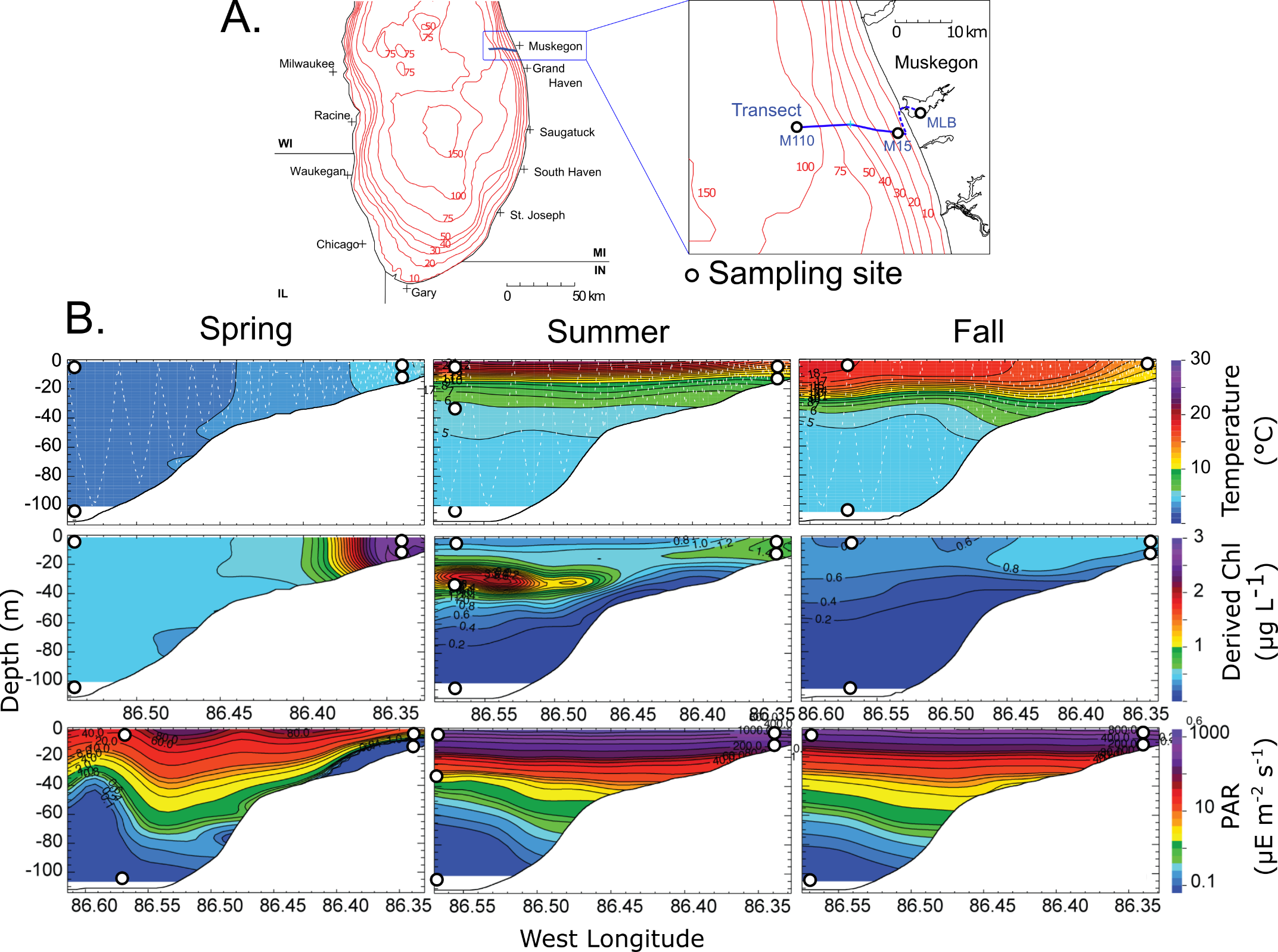
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Query** | **Subject** | **Organism** | **Total Score** | **Query Cover** | **E value** | **Percent Identity** |
| 2757589073 (CheA) | [RLT54355.1](https://www.ncbi.nlm.nih.gov/protein/RLT54355.1?report=genbank&log$=prottop&blast_rank=1&RID=ZTXKH9H9016) (CheA) | Chloroflexi bacterium | 696 | 96% | 0.0 | 56.18% |
| 2757589077 (CheB) | [RLT47469.1](https://www.ncbi.nlm.nih.gov/protein/RLT47469.1?report=genbank&log$=prottop&blast_rank=1&RID=ZTXKH9H9016) (CheB) | Chloroflexi bacterium | 464 | 97% | 3e-161 | 64.96% |
| 2757589076 (CheR) | [RLT47470.1](https://www.ncbi.nlm.nih.gov/protein/RLT47470.1?report=genbank&log$=prottop&blast_rank=1&RID=ZTXKH9H9016) (CheR) | Chloroflexi bacterium | 266 | 94% | 5e-85 | 49.08% |
| 2757589073 (CheA) | WP\_062405381.1 (CheA) | *Limnohabitans* sp. 63ED37-2 | 258 | 46% | 1.00E-74 | 40.36% |
| 2757589077  (CheB) | WP\_082431645.1 (CheB) | *Limnohabitans* sp. 63ED37-2 | 171 | 93% | 5.00E-51 | 36.00% |
| 2757589076  (CheR) | ALK88970.1 (Chemotaxis protein methyltransferase Cher2) | *Limnohabitans* sp. 63ED37-2 | 67.8 | 60% | 7.00E-14 | 29.15% |
| 2757589076  (CheR) | 2757587174 (CheR) | MAG7.SU-MLB-SD (LimC) | 82 | 72% | 7e-22 | 29.13% |
| 2757589077  (CheB) | 2757587175 (CheB) | MAG7.SU-MLB-SD (LimC) | 164 | 93% | 2e-51 | 35.07% |

**Supplementary Table S3**: Number of inferred strains by DESMAN in 10 replicate runs which were performed for a potential number of strains ranging between 2 and 12. The number of assessed strains in the best run are provided, as well as the number of strains that could be confidently resolved from that run. The average error on these inferences and the number of variant positions for resolving the strains are provided as well.

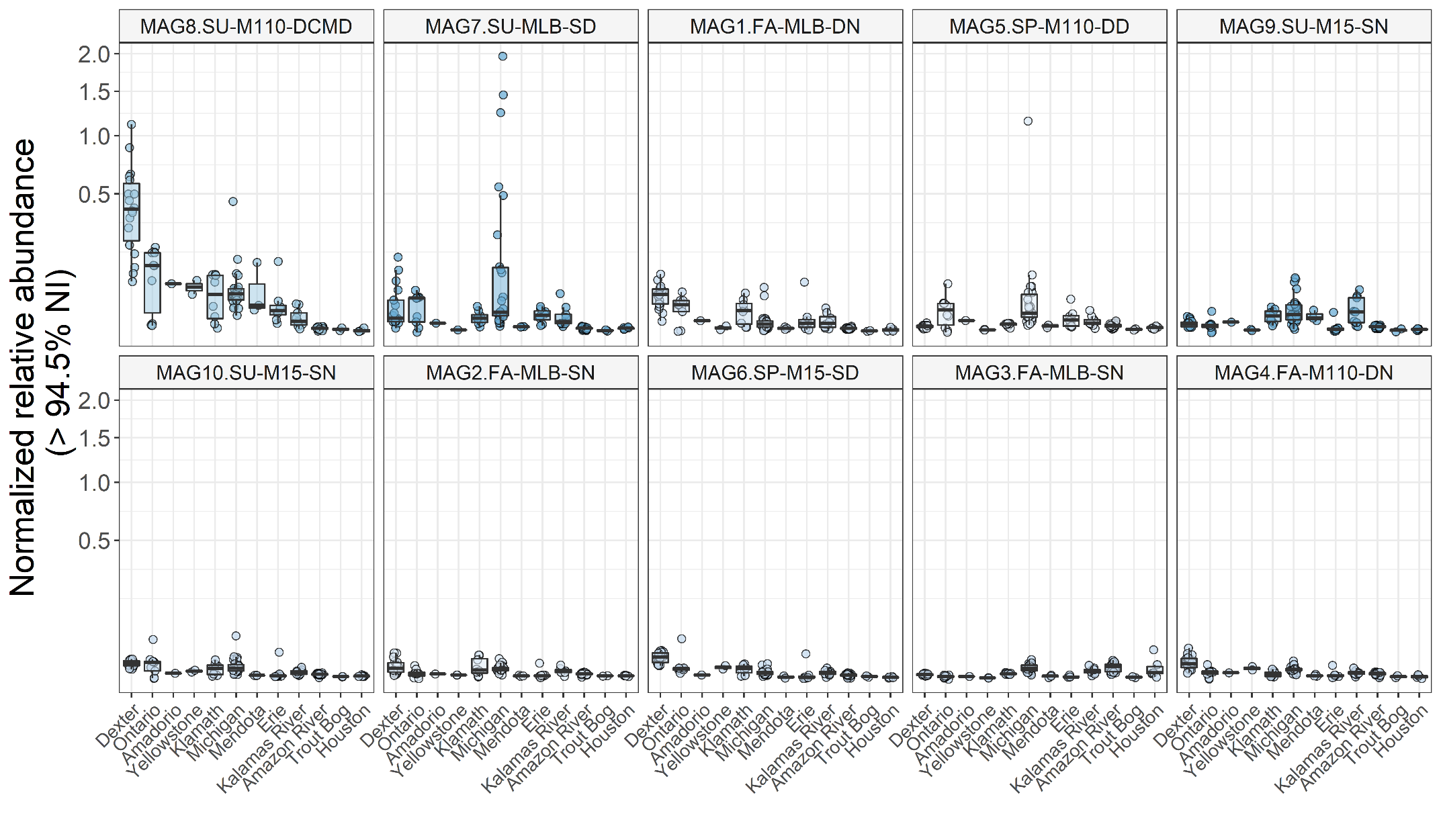
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Limnohabitans* MAGs | Number of strains in best DESMAN run (n = 10) | Number of resolved strains | Average error (n = 10) | Number of variant positions |
| MAG1.FA-MLB-DN | 2 | 2 | 0.025 | 8 |
| MAG2.FA-MLB-SN | 4 | 3 | 0.043 | 12 |
| MAG3.FA-MLB-SN | 2 | 2 | 0.021 | 8 |
| MAG4.FA-M110-DN | 3 | 3 | 0.036 | 12 |
| MAG5.SP-M110-DD | 6 | 3 | 0.056 | 12 |
| MAG6.SP-M15-SD | 2 | 2 | 0.013 | 8 |
| MAG7.SU-MLB-SD | 2 | 2 | 0.021 | 8 |
| MAG8.SU-M110-DCMD | 11 | 6 | 0.043 | 24 |
| MAG9.SU-M15-SN | 7 | 3 | 0.028 | 12 |
| MAG10.SU-M15-SN | 6 | 3 | 0.032 | 12 |

**Supplementary Table S4**: JGI sequencing project identifiers for the raw metagenomic and metatranscriptomic data.

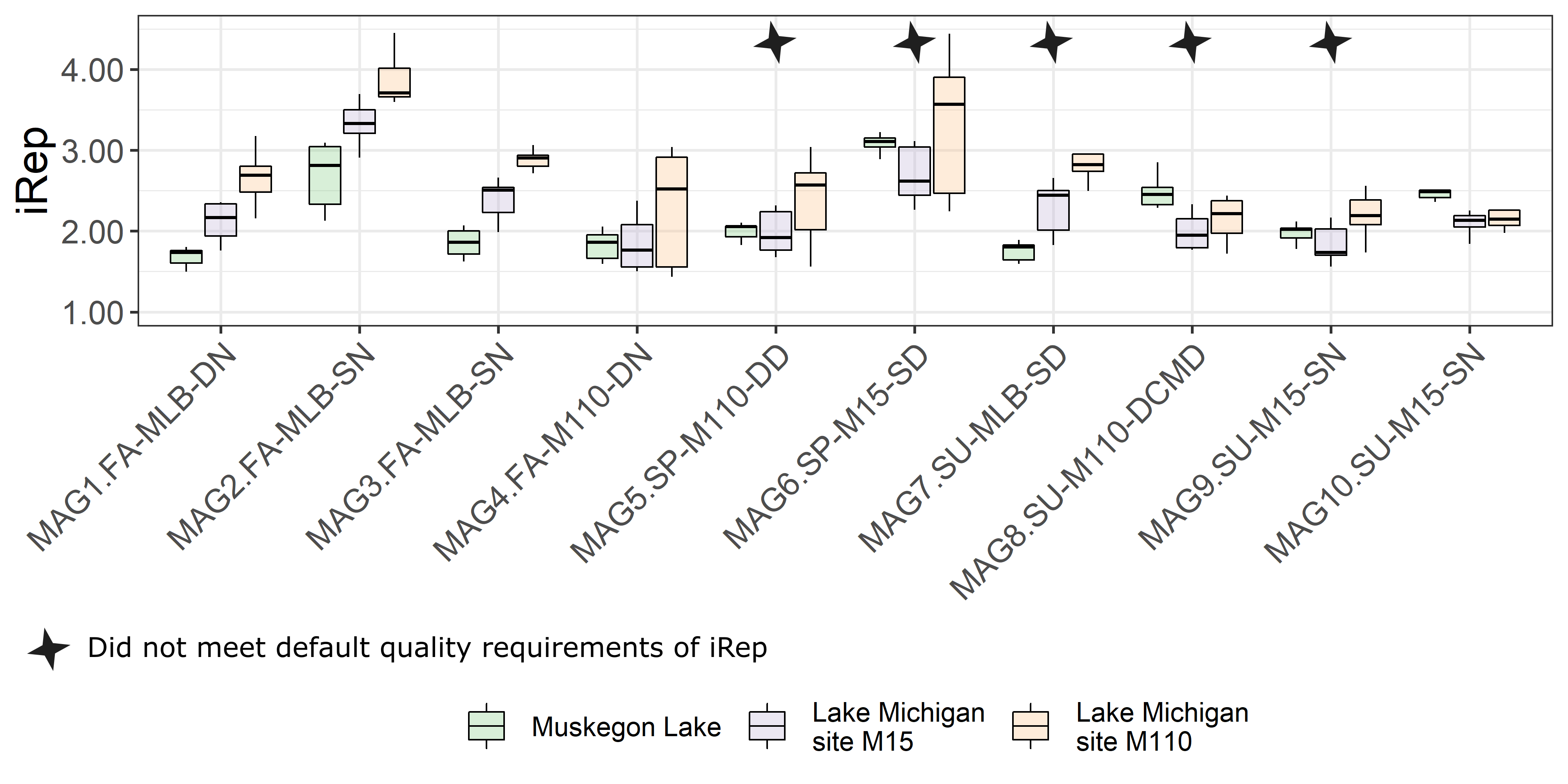
|  |  |  |
| --- | --- | --- |
| Sample name | Project ID metatranscriptomics | Project ID metagenomics |
| Sp13.BD.MM15.SN | 1035007 | 1035079 |
| Sp13.BD.MM15.SD | 1035010 | 1035082 |
| Sp13.BD.MM15.DD | 1035013 | 1035085 |
| Sp13.BD.MM110.SN | 1035016 | 1035088 |
| Sp13.BD.MM110.SD | 1035019 | 1035091 |
| Sp13.BD.MM110.DD | 1035022 | 1035094 |
| Sp13.BD.MLB.SN | 1035025 | 1035097 |
| Su13.BD.MM15.SN | 1035028 | 1035100 |
| Su13.BD.MM15.SD | 1035031 | 1035103 |
| Su13.BD.MM15.DN | 1035034 | 1035106 |
| Su13.BD.MM110.SN | 1035037 | 1035109 |
| Su13.BD.MM110.SD | 1035040 | 1035112 |
| Su13.BD.MM110.DN | 1035043 | 1035115 |
| Su13.BD.MM110.DCMD | 1035046 | 1035118 |
| Su13.BD.MLB.SD | 1035049 | 1035121 |
| Su13.BD.MLB.DD | 1035052 | 1035124 |
| Fa13.BD.MM15.SN | 1035055 | 1035127 |
| Fa13.BD.MM15.SD | 1035058 | 1035130 |
| Fa13.BD.MM15.DN | 1035061 | 1035133 |
| Fa13.BD.MM110.SN | 1035064 | 1035136 |
| Fa13.BD.MM110.SD | 1035067 | 1035139 |
| Fa13.BD.MM110.DN | 1035070 | 1035142 |
| Fa13.BD.MLB.SN | 1035073 | 1035145 |
| Fa13.BD.MLB.DN | 1035076 | 1035148 |



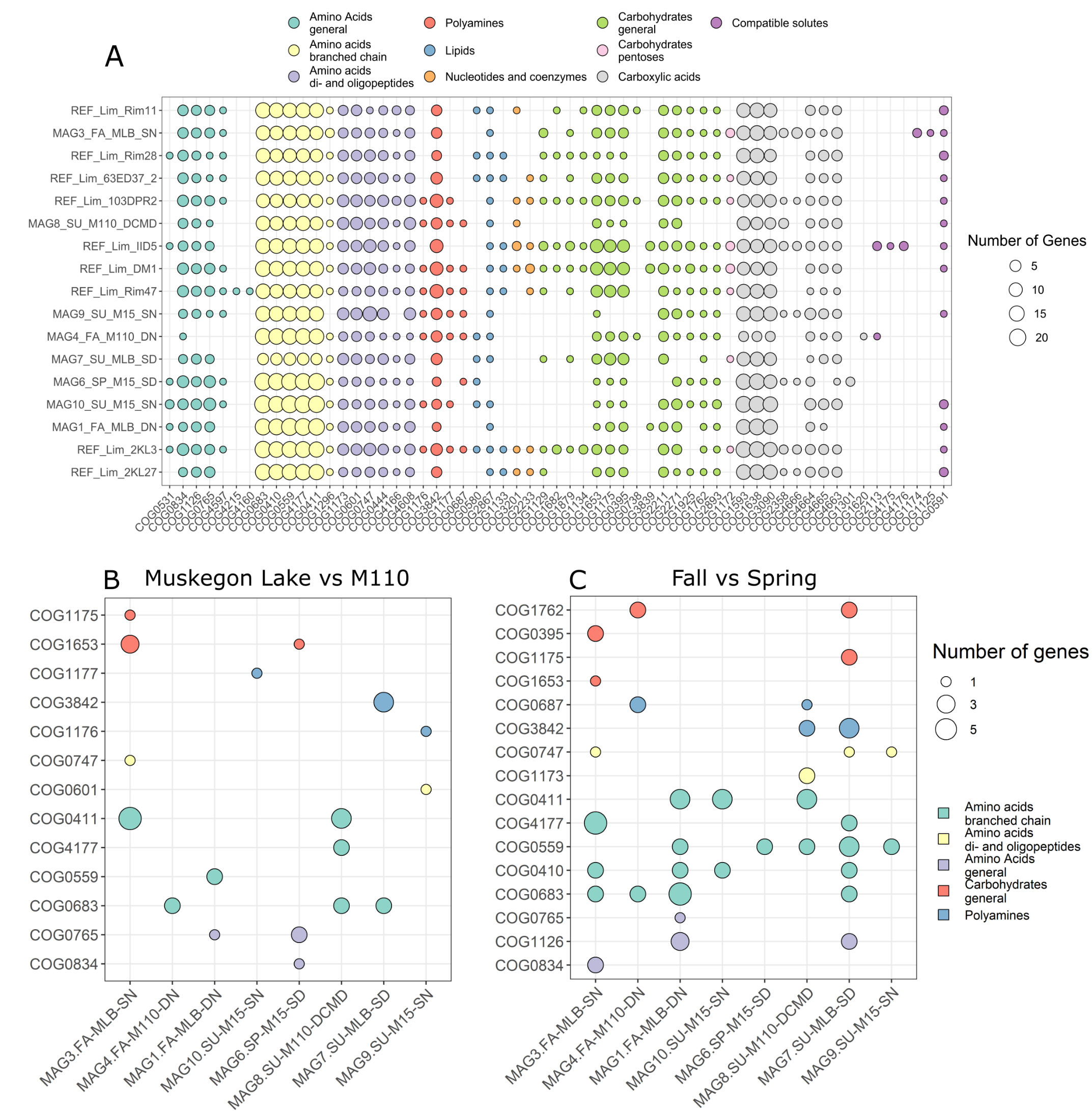
**Supplementary Figure 1**: **A**. Sampling transect across Muskegon Lake and Lake Michigan with depth contours and sampling sites (circle symbols; M110 (offshore), M15 (nearshore), MLB (Muskegon Lake) highlighted. At each location samples were taken at various depths, seasons and time of day. **B**. Temperature, chlorophyll a and photo active radiation (PAR) profiles between M110 and M15 sites as determined by the plankton survey system (dotted lines indicate the travel path). Circles highlight sampling sites along the Lake Michigan transect. Figures were adapted from (1).



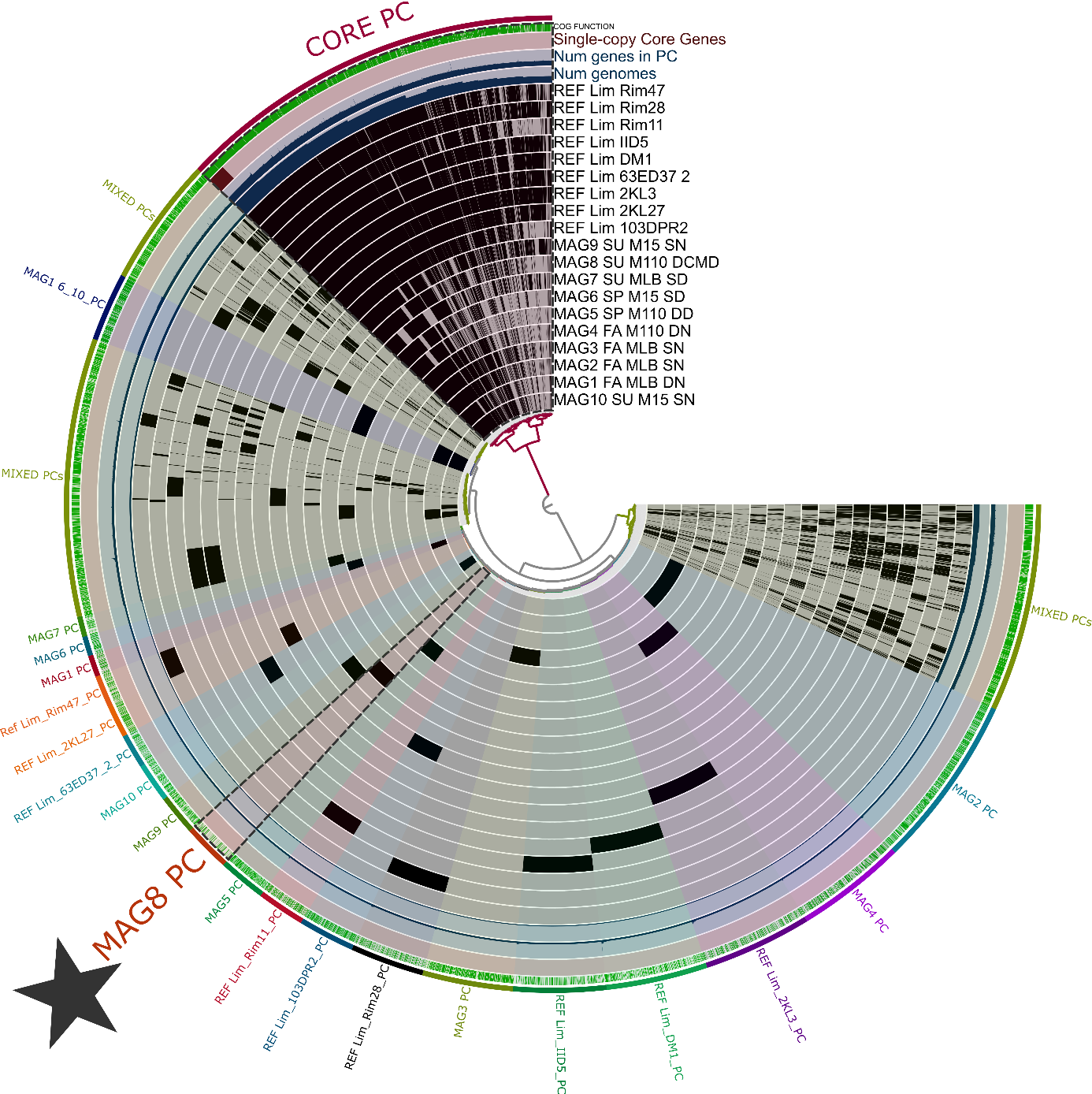
**Supplementary Figure 2**: Inferred relative abundances of closely related populations (SDPs) to the 10 putative Limnohabitans MAGs reconstructed from Lake Michigan metagenomic data. One million reads were competitively mapped to the 10 putative Limnohabitans MAGs. Abundances were normalized for MAG sizes. Further information regarding the sampled environments can be retrieved from Neuenschwander et al. (2017).



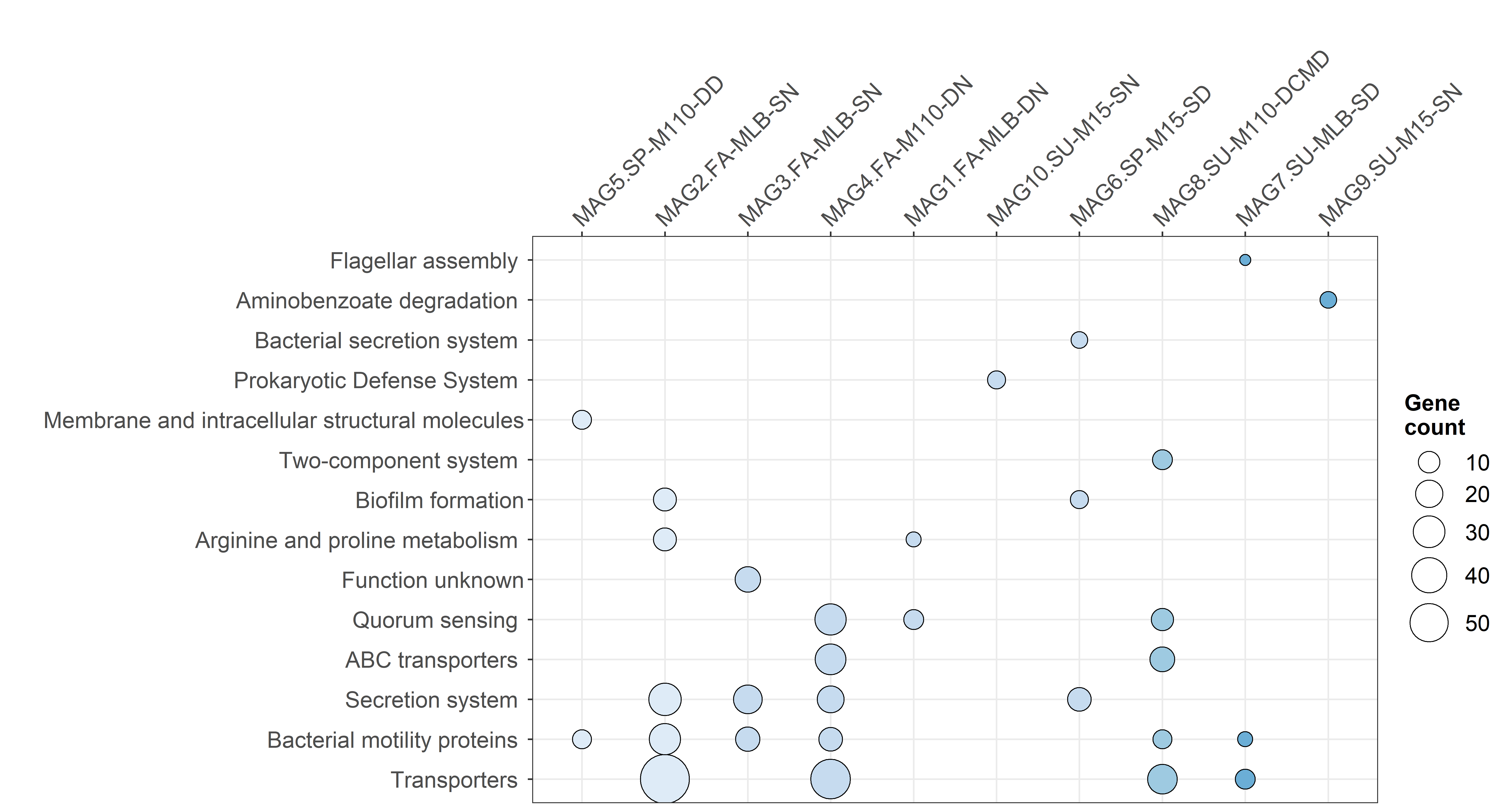
**Supplementary Figure 3**: Boxplots of index of replication (iRep) of each putative Limnohabitans MAG at each sampling station for the spring, summer, and fall seasons and at the various depths.



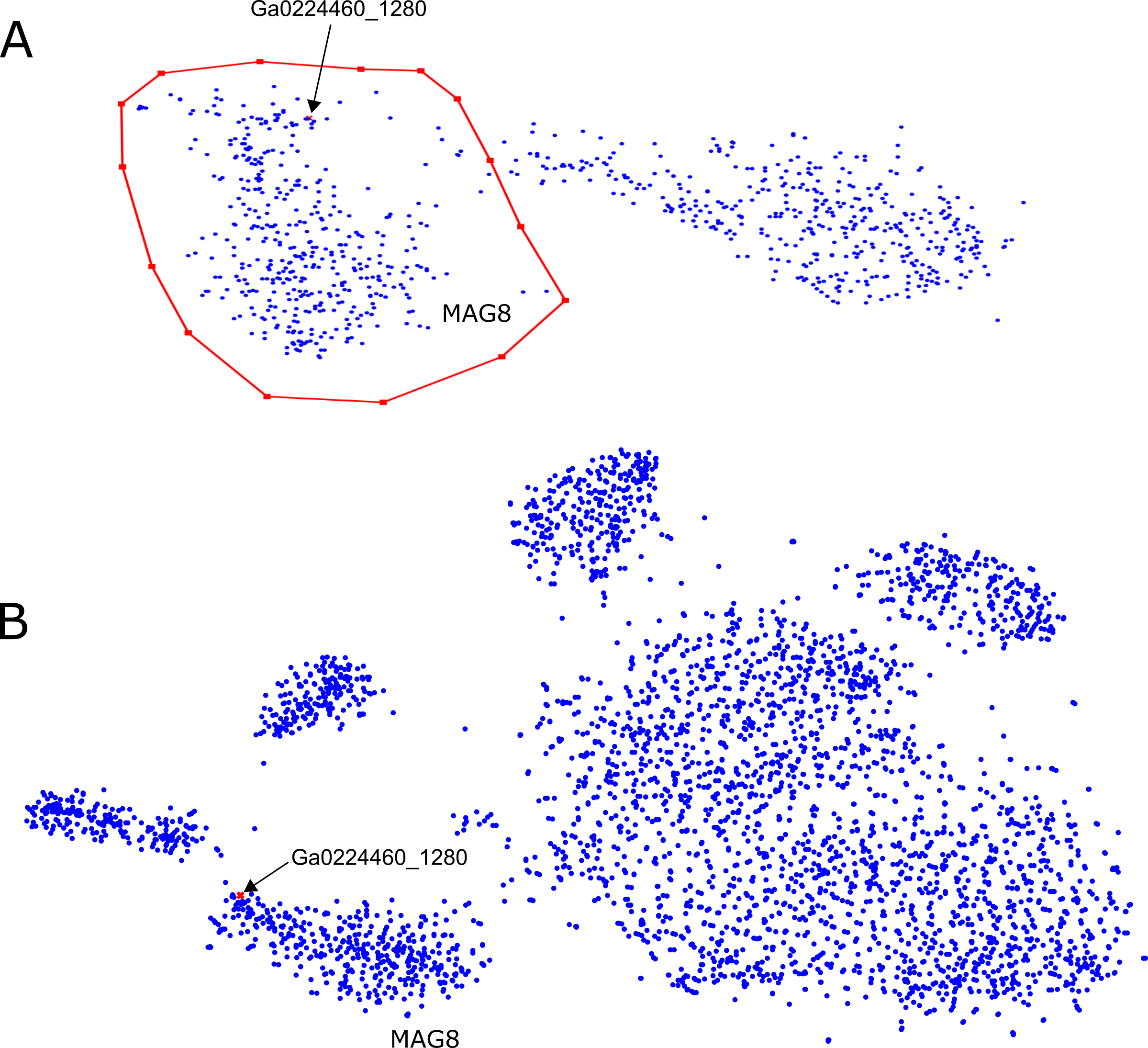
**Supplementary Figure 4**: A. *Frequencies of dissolved organic carbon (DOC) transporter genes in the Limnohabitans MAGs and available isolate genomes (2). MAGs are ordered according to hierarchical clustering of their DOC frequency profiles. B. Differentially expressed genes between Muskegon Lake and Lake Michigan M110 site controlled for season. C. Differentially expressed genes between the Fall and Spring season controlled for site (C).*



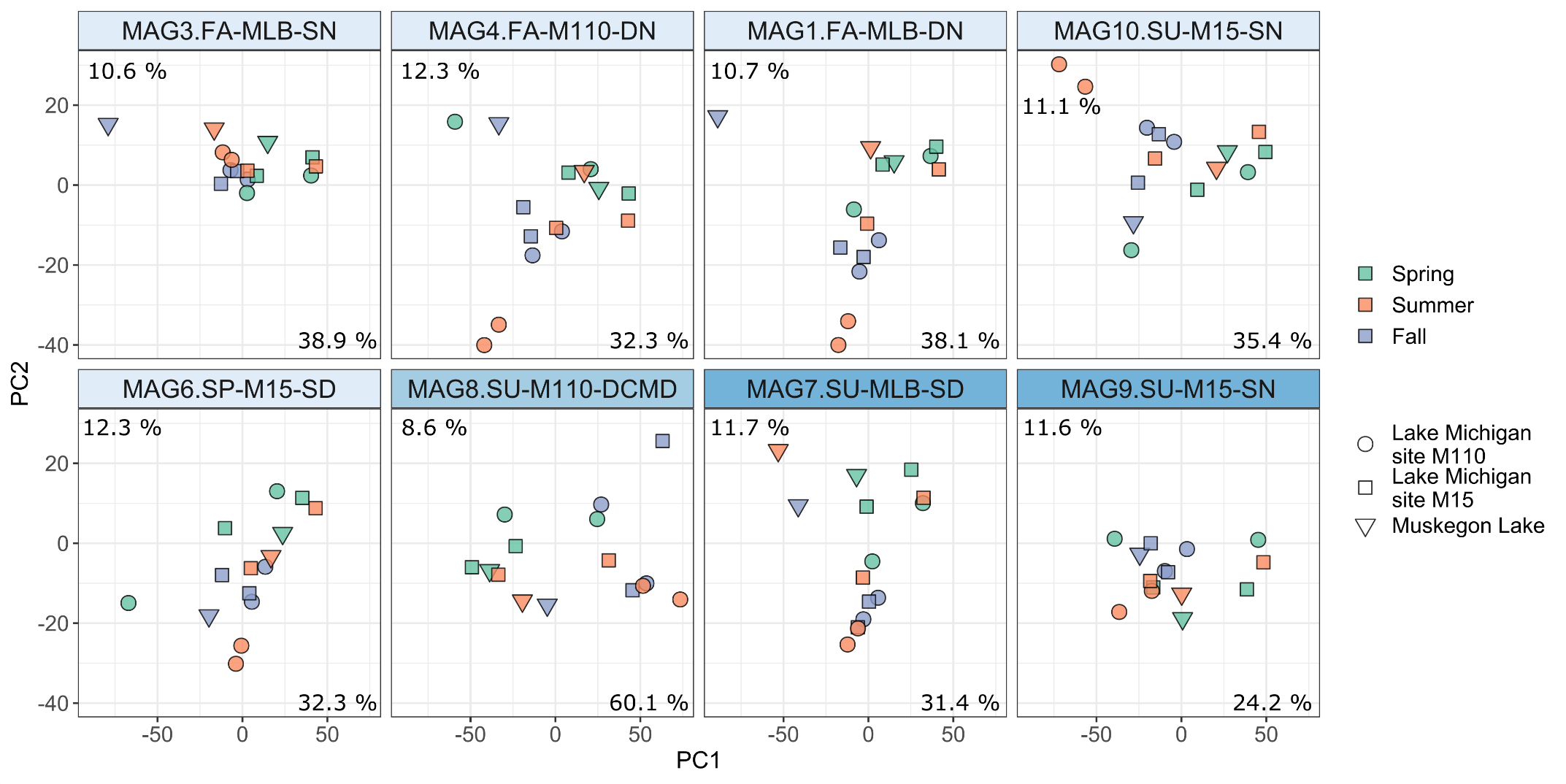
**Supplementary Figure 5**: Pangenome analysis of Limnohabitans genomes and MAGs. The star indicates the accessory genome of MAG8.SU-M110-DCMD. Anvi’o database is available at <https://doi.org/10.6084/m9.figshare.7547159>.

**

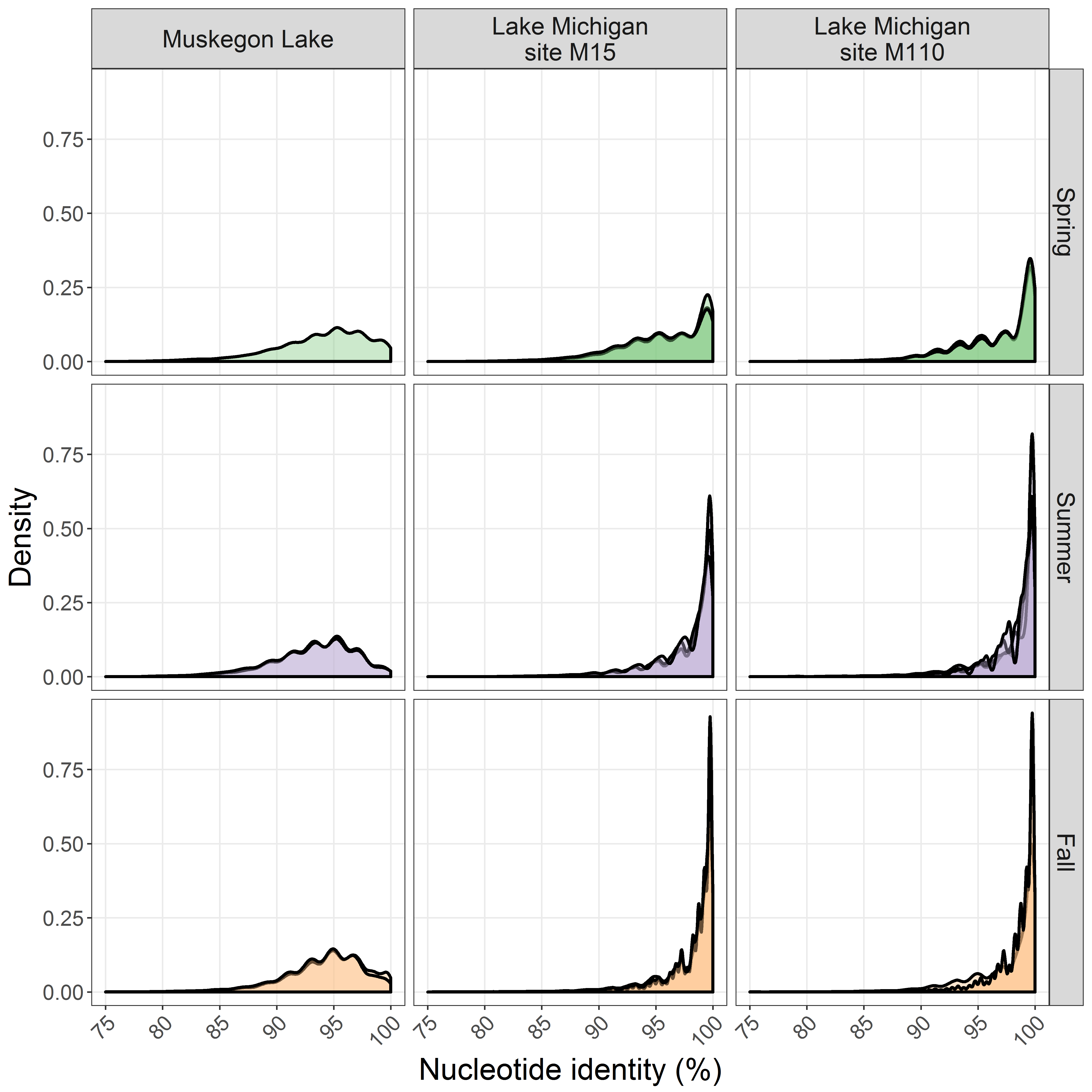
**Supplementary Figure 6**: Heatmap of KEGG subsystem categories significantly enriched (adj. p-value < 0.05) in the accessory genomes of each putative Limnohabitans MAG.



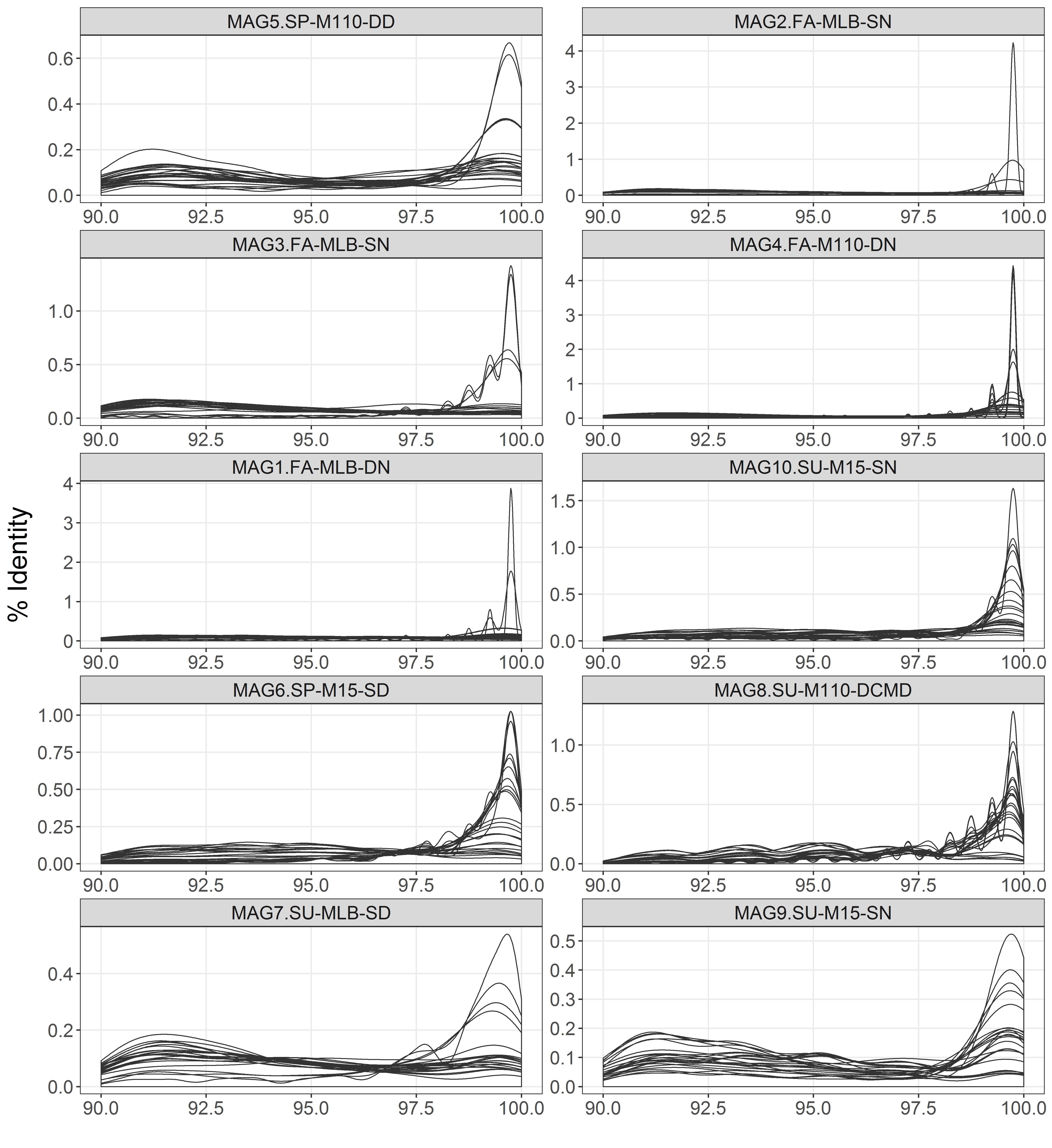
**Supplementary Figure 7**: **A**. Snapshot of the binning of MAG8.SU-M110-DCMD using Vizbin (k = 5, default settings) in the original metagenomic assembly with the contig containing the chemotaxis genes highlighted (Ga0224460\_1280). **B**. Vizbin tSNE plot of the contigs of all final dereplicated putative Limnohabitans MAGs. The contig containing the chemotaxis genes is highlighted (Ga0224460\_1280).



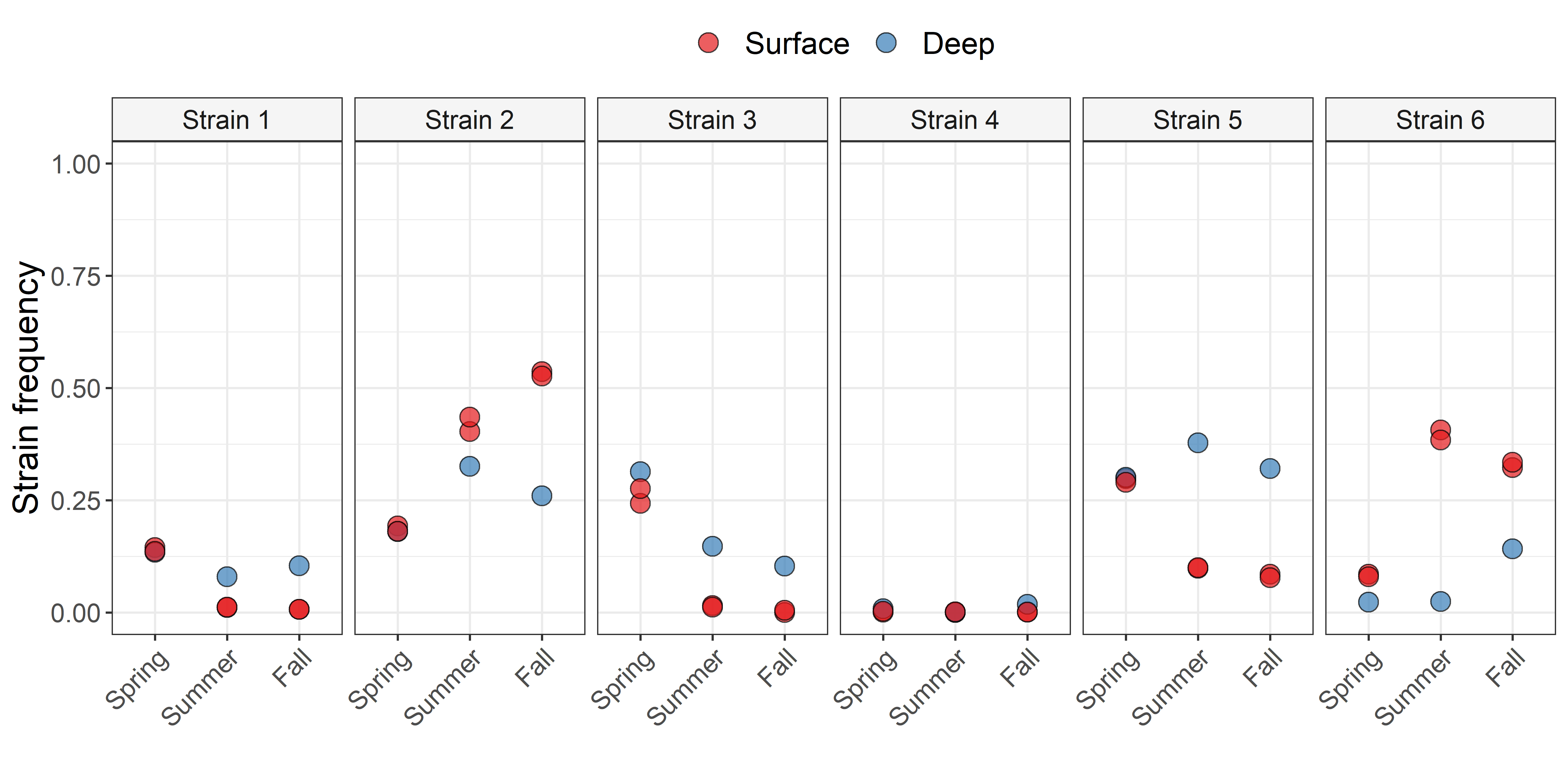
**Supplementary Figure 8**: *Principal component analysis of TMM-normalized expression profiles of MAGs. Variation explained by each PC is given next to the respective axis.*



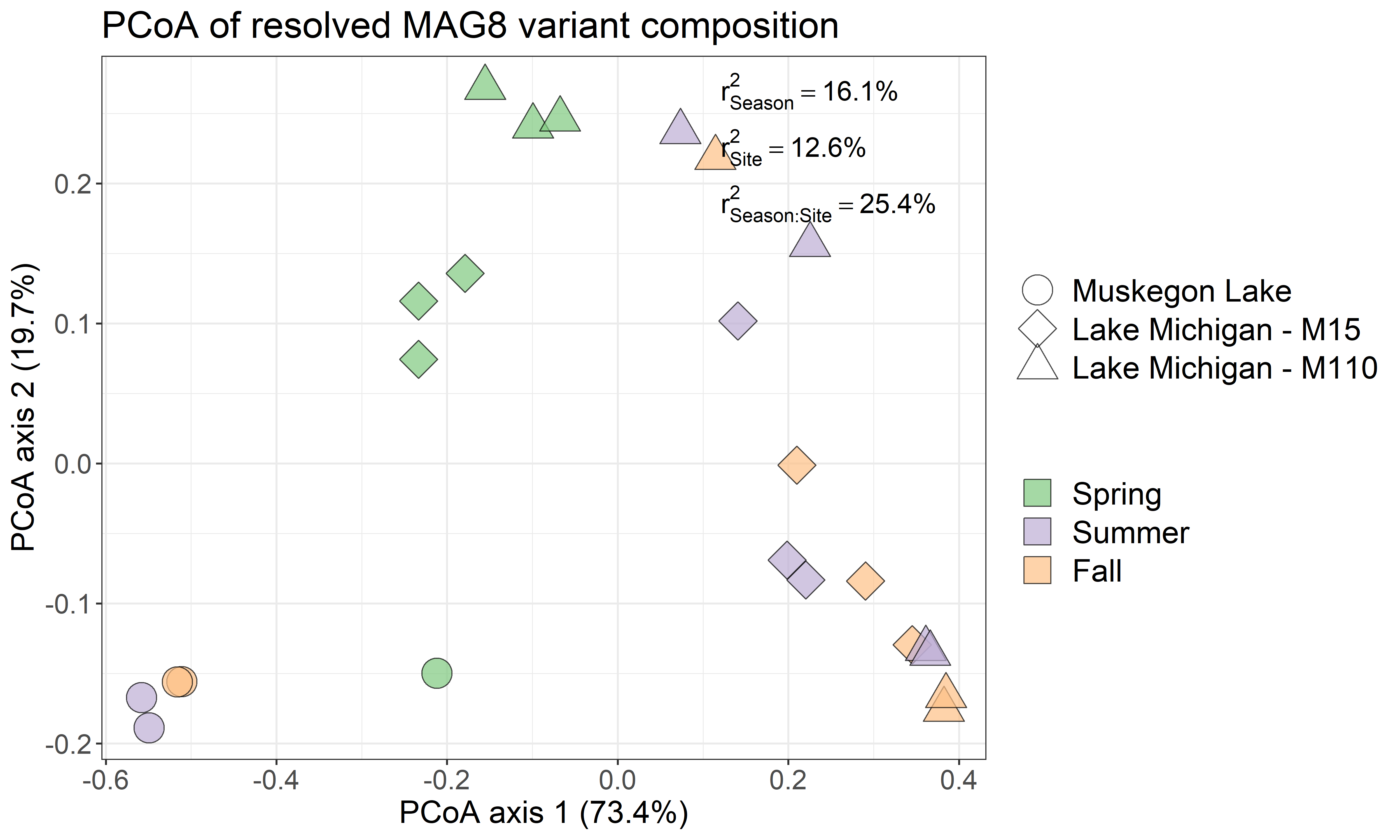
**Supplementary Figure 9**: Nucleotide identity distribution of metagenomic reads mapping to MAG8.SU-M110-DCMD after competitive recruitment of one million reads per sample. Density estimation was made using the bw.nrd0 rule-of-thumb bandwidth.



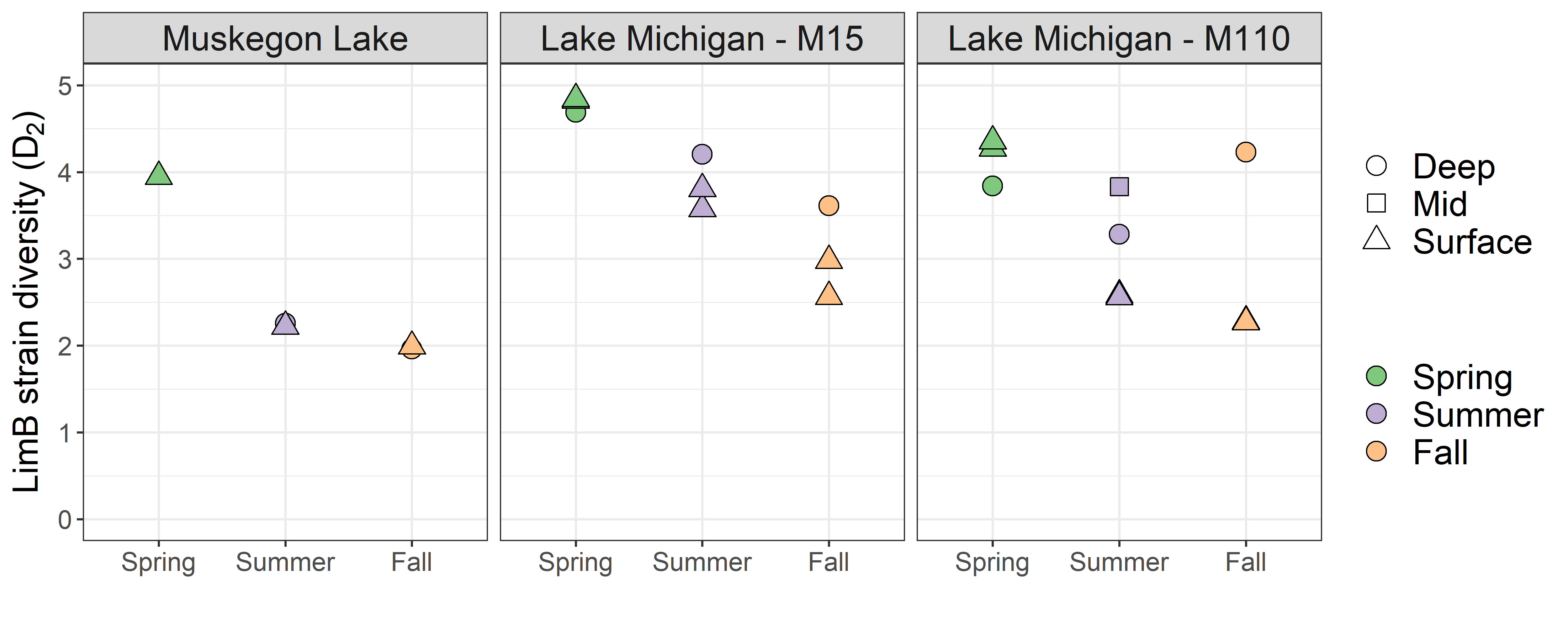
**Supplementary Figure 10**: Nucleotide identity distribution of metagenomic read recruitment to the various putative Limnohabitans MAGs for all samples. Density estimations used Gaussian kernels and bandwidths were estimated using the “nrd0” estimator.



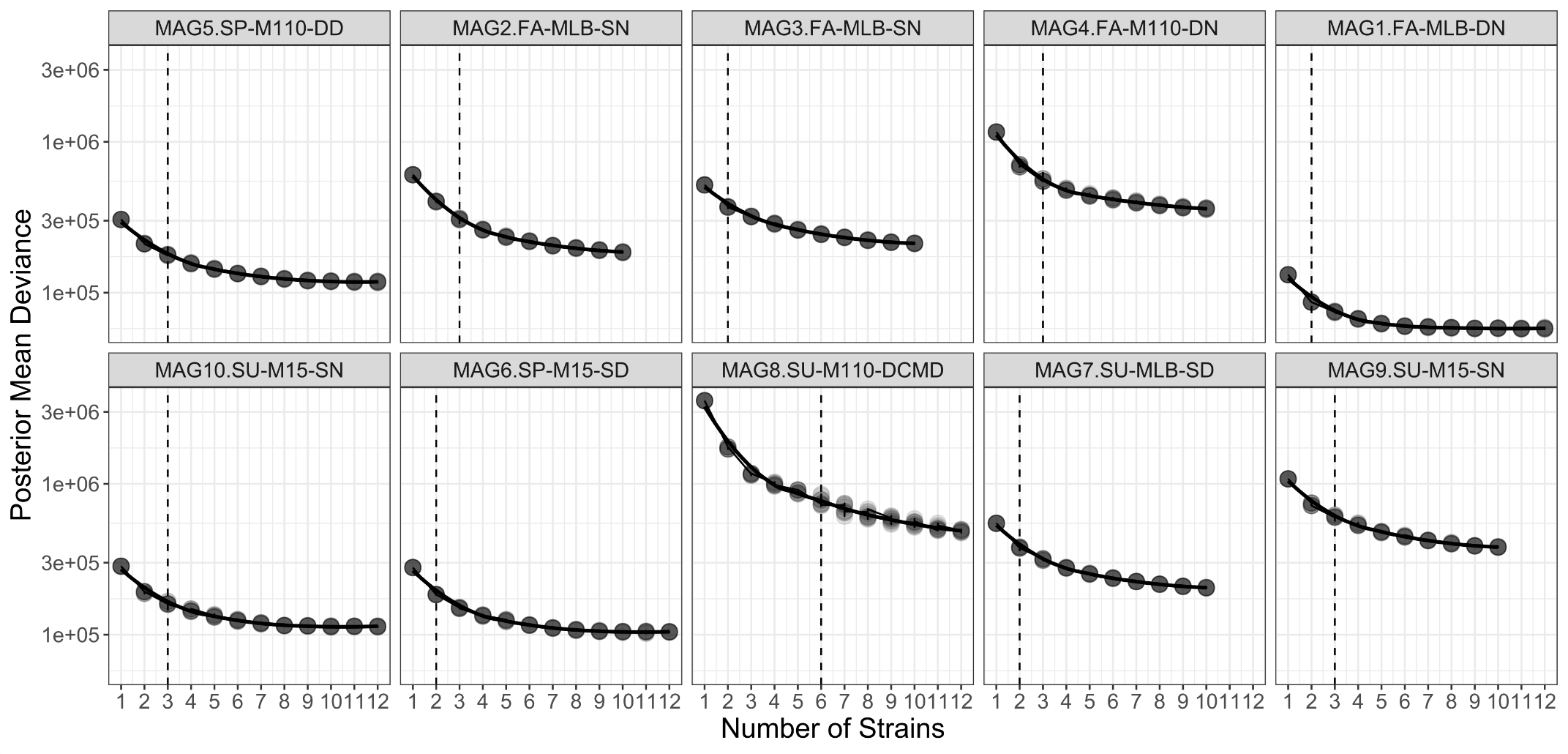
**Supplementary Figure 11**: MAG8.SU-M110-DCMD strain frequencies at the bottom and surface of the offshore station (M110) in Lake Michigan.



**Supplementary Figure 12**: Principal Coordinate Analysis (PCoA) of the strain frequencies of MAG8.SU-M110-DCMD. The variance explained by spatiotemporal categories as inferred by PERMANOVA is overlaid on the plot.



**Supplementary Figure 13**: MAG8 strain diversity as inferred by the Hill diversity of order 2 (Inverse Simpson index) across the different seasons, sites and depths.



**Supplementary Figure 14**: Posterior Mean Deviance (proxy for model fit) for the DESMAN runs at different preset number of strains. For each number of strains, 10 DESMAN runs were conducted. The dotted line indicates the number of strains for which the best DESMAN run was found (resolvenhap.py script). Additional number of strains beyond this point did not significantly increase the model fit.

1. Denef VJ, Mueller RS, Chiang E, Liebig JR, Vanderploeg HA.2016. Chloroflexi CL500-11 Populations That Predominate Deep-Lake Hypolimnion Bacterioplankton Rely on Nitrogen-Rich Dissolved Organic Matter Metabolism and C-1 Compound Oxidation. Applied and Environmental Microbiology 82:1423-1432.

2. Poretsky RS, Sun S, Mou X, Moran MA.2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environmental Microbiology 12:616-627.